A Statistical Evaluation of the Collaborative Convective Forecast Product (CCFP) and the Convective Significant Weather Advisory (C-SIGMET): 2001 and 2002

4 August 2002

Minor revisions: 13 August 2002

Jennifer Luppens Mahoney[1*], Barbara G. Brown[2], Joan E. Hart[1,3], and Judy Henderson[1]

---

* Corresponding authors address: 325 Broadway, Boulder, CO 80305-3328, mahoney@fsl.noaa.gov
[1] NOAA Research-Forecast Systems Laboratory
[2] National Center for Atmospheric Research
[3] Joint collaboration with the Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, CO.

1. Introduction

This report summarizes the results available from the 2001 and 2002 evaluation of the forecast capability of the National Weather Service (NWS)-issued Convective Significant Meteorological Advisories (C-SIGMET) and the Collaborative Convective Forecast Product (CCFP). This study was undertaken by the Aviation Forecasts and Quality Assessment Product Development Team of the Federal Aviation Administration (FAA) Aviation Weather Research Program (AWRP).

The purposes of the evaluations were to i) develop a baseline for the quality of the C-SIGMETs and the CCFP, ii) to demonstrate to-date progress in the improvement of the forecasts, iii) examine the similarities and differences between the two forecasts, and iv) perform an evaluation that is independent, consistent, comprehensive, and fair. To meet the first two goals, verification methods, established in 1999 (Mahoney et al. 2000a), have enabled the consistent generation of statistics for the C-SIGMETs and the CCFP. This consistency has led to the development of a baseline that is being used to show improvements in skill in the two forecasts. To meet the third goal, the differences between 2-h C-SIGMET and CCFP forecasts will be discussed. Finally, the fourth goal was met by enabling the independent evaluation of the results by the Quality Assessment Group (QAG), which is a collaborative effort between the verification groups of the NOAA Forecast System Laboratory (FSL) and the National Center for Atmospheric Research (NCAR) Research Applications Program (RAP).

This report is one in a series that describe the quality and skill of the C-SIGMET and CCFP forecasts. The previous verification studies focused on the development and testing of the verification methods as well as understanding the attributes of the forecasts (Mahoney et al. 2000a, 2000b, 2001, 2002a, 2002b). The results presented take advantage of the use of consistent verification methods to establish a baseline for the C-SIGMETs, and are used to describe improvements in forecast skill since 2000.

The focus of this third evaluation is the period from 1 April–31 October 2001, although comparisons with results for the summer of 2000 and 2002 will also be presented. Ongoing verification of the C-SIGMET and CCFP forecasts are available through the Real-Time Verification System (RTVS; Mahoney et al. 1997, 2002b) Web site at http://www-ad.fsl.noaa.gov/fvb/rtvs/; link convective or CCFP. The displays and analyses presented on RTVS will be summarized in this report.

The forecasts and observations are considered in Sections 2 and 3, respectively. The evaluation approach is described in Section 4, and the methodology used to generate the forecast/observation pairs and the statistical verification approaches are discussed in Section 5. Results are presented in Section 6. Finally, a summary and some conclusions are presented in Section 7.

2. Approach

The C-SIGMET and CCFP forecasts issued by the Aviation Weather Center (AWC) were included in this study. All forecast issue and lead times were evaluated. The forecasts were evaluated using the National Convective Weather Detection (NCWD) product developed by the NCAR/RAP. The evaluation began on 1 April and continued through 31 October 2001. Traditional statistical tables and plots were generated through RTVS and were provided to users through the Web-based interface. In addition, graphical displays, such as those shown in Figs. 1 a and b, were provided through the Web-based interface for each forecast issue and lead

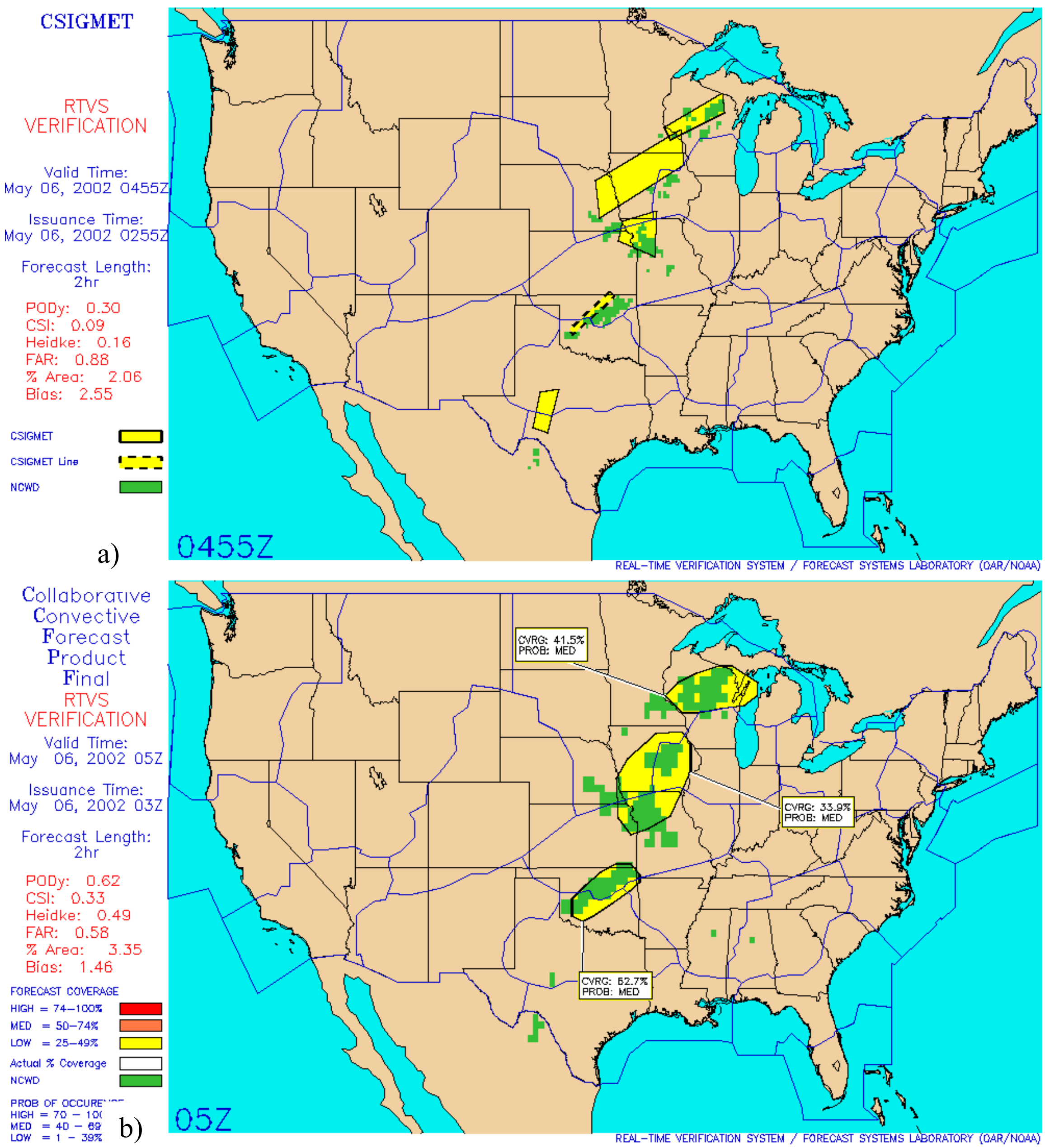


Fig. 1 a and b. Example verification maps for (a) the C-SIGMETs and (b) the CCFP, for forecasts 2-h lead issued at 0300 UTC on 6 May 2002. Polygons represent forecasts and blocks represent NCWD observations. Statistics computed for each issue/lead time are presented left of the display.

time for the C-SIGMETs (Fig. 1a) and CCFP (Fig. 1b). The maps may also be viewed in animation from the Web site. Verification statistics were computed for all forecast issue and lead times for the CCFP and for the 0-, 1-, and 2-h forecast lead times from the hourly issuances for the C-SIGMETs.

The forecasts and observations were processed by RTVS as they became available to the system. If data were missing or were late getting to the system, and/or the system processing or data transmission failed, results were not generated for that specific time period in near real time. However, once the evaluation was completed, attempts were made to fill in missing time periods and reanalyze the data.

3. Forecasts

The convective forecasts considered in the evaluation are described in the following paragraphs.

Collaborative Convective Forecast Product (CCFP): The CCFP is prepared through a multistep collaborative process (Phaneuf and Nestoros 1999; Hudson and Foss 2002) that begins with AWC forecasters, but includes participation from airline meteorologists and dispatchers, as well as meteorologists from the Center Weather Service Units (CWSUs) at the Air Route Traffic Control Centers (ARTCCs). The CCFP is used as a strategic decision aid by the decision-makers at the airlines and the Air Traffic Control System Command Center (ATCSCC) for rerouting air traffic around convective weather. The issue and lead times produced during 2001 for the CCFP are summarized in Table 1 and the forecast criteria are presented in Table 2. The CCFP text-formatted file, obtained

Table 1. Issue, leads and valid times for the CCFP

Issue Time (UTC)	*Lead Time (h)*	*Valid Time (UTC)*
0300	2, 4, 6	0500, 0700, 0900
*0700	2, 4, 6	0900, 1100, 1300
*1100	2, 4, 6	1300, 1500, 1700
1500	2, 4, 6	1700, 1900, 2100
1900	2, 4, 6	2100, 2300, 0100
2300	2, 4, 6	0100, 0300, 0500

*These forecasts available from 18 June–31 October 2001.

Table 2. Forecast criteria for the CCFP

Cloud tops	*Coverage*	*Probability*	*Movement*	*Growth*
Minimum 25,000 ft	Minimum 25%	Low < 40%	Movement indicated by arrows	Fast positive (++)
25,000-31,000 ft	Low 25-49%	Medium 40-69%		Positive (+)
31,000-37,000 ft	Medium 50-75%	High > 70%		No change (NC)
> 37,000 ft	High > 75%			Negative (-)

from the National Centers for Environmental Prediction (NCEP), is decoded and reformatted before it is used by RTVS. Only the results for the final CCFP forecasts are presented in this report.

Convective Significant Meteorological Advisory (C-SIGMET) – The C-SIGMET is an operational forecast issued hourly by AWC for thunderstorms and related phenomena which imply the associated occurrence of turbulence, icing, and convective low-level wind shear. The specific criteria for issuing a C-SIGMET are summarized by NWS (1991). The C-SIGMET polygon is valid at the time it is issued. The 1- and 2-h C-SIGMET forecasts are derived by moving the polygon as stated by the speed and direction vector listed in the body of the forecast (Foss, personal communication). The text-formatted file, obtained from NOAAPORT, is decoded and reformatted before being used in RTVS.

4. Observations

The NCWD observations were used to evaluate the forecasts. In earlier evaluations, lightning and radar observations were used separately to evaluate the convective forecasts. The results from those evaluations indicated that better overall observations of convective weather were provided by the NCWD rather than from lightning or radar observations alone (Mahoney et al. 2000a).

The NCWD (Mueller et al. 1999), developed by the FAA's Convective Weather Product Development Team (Sankey et al. 1997, Kulesa et al. 2002) was obtained on a 4-km grid in near real time from the AWC and was used to evaluate the forecasts. The NCWD is a convective hazard detection field depicting areas of convective weather that may be hazardous to aviation. The hazard field is based on WSR-88D national radar mosaics and cloud-to-ground lightning observations from the National Lightning Detection Network (Orville 1991). The Vertically Integrated Liquid Water (VIL) data used in NCWD provides information about the intensity of a storm throughout its vertical extent and provide a proxy for vertical development. A VIL threshold of 3.5 kg m^{-2} (Level 3) and/or a lightning rate of at least 3-5 strikes in 10 minutes were used to delineate storms for the verification analyses. Additional information describing the NCWD can be obtained at http://cdm.aviationweather.noaa.gov/ncwf/ncwf_wt/.

5. Verification Methodology

The methodologies obtained from tests performed during the 2000 evaluation (Mahoney et al. 2001) of the C-SIGMET and CCFP were applied here and are described in this section.

5.1. Mechanics

5.1.1 Converting the forecasts and observations to a common grid

Prior to computing the statistics, the forecasts and observations were converted to a common grid, which was chosen to represent the scale of the forecast. Mahoney et al. (2001) found that a 40-km grid best represents the scale of the CCFP, and a 20-km grid best represents the scale of the C-SIGMET. The technique for mapping the forecasts to the common grid included labeling the grid box on the common grid with a *Yes* forecast when any part of the forecast polygon intersected the grid box. In some places, this technique blossomed the size of the forecast area, but not substantially. If the forecast polygon did not intersect a grid box, then a *No* forecast was assigned to that box. Each forecast for every issue and lead-time was mapped to a common grid.

A similar procedure was applied when mapping the 4-km NCWD observations to the common grid (i.e., 40-km for the CCFP and 20-km for the C-SIGMETs). This procedure was more difficult to define than that used for the forecasts because the scales of the forecasts and observations are largely different. After testing different filtering techniques (Mahoney et al. 2000a, 2001), the criterion that worked best for mapping the 4-km NCWD to the common grid involved labeling a grid box on the common grid to be a *Yes* observation when at least one 4-km NCWD box with a VIL value greater than 3.5 fell within the larger grid box. Again, this procedure can have the effect of "blooming" the size of an observed area. However, this approach is reasonable when considering (a) the amount of regulated buffer required between aircraft and convection (nominally, 20 nautical miles), and (b) the NCWD threshold that is applied implies that even a single 4-km grid-box is likely to be embedded in a larger (though less intense) convective region.

The problems with this approach include the ability to directly compare the CCFP and the C-SIGMETs because of the differing scales used to represent the forecast grid, the "sufficient extent" criteria of 3,000 sq. miles for the C-SIGMET is excluded from the current verification approach, and the procedure for assigning the convection to a grid not only has the effect of blooming the convective activity, but may also include convection that would have been excluded from the forecast because it did not meet the size criteria. ***Nevertheless, it's the consistency with which the verification methods are applied over time and the relative comparison between the statistical results that allow detection of improvements in the forecast skill.***

5.1.2 Matching forecasts to observations

Once the forecasts and observations were mapped to a common grid, the forecast grid was matched to the observation grid, producing forecast/observation pairs. Specifically, each box on the observation grid was matched to the corresponding box on the forecast grid to create a forecast/observation pair at each grid box. For example, a *Yes* forecast box and a *Yes* observation box produced a *Yes-Yes* forecast/observation pair. Similarly, a *Yes* forecast and a *No* observation produced a *Yes-No* pair, and so on, filling in the four cells of the statistical contingency table as shown in Table 3. In the case of the CCFP, each forecast area, regardless of its predicted coverage, was considered a forecast of convection and was matched to the observations. The actual coverage for the CCFP is computed in a separate analysis described later.

Table 3. Basic contingency table for evaluation of dichotomous (e.g., *Yes/No*) forecasts. Elements in the cells are the counts of forecast-observation pairs.

Forecast	***Observation***		***Total***
	Yes	***No***	
Yes	YY	YN	YY+YN
No	NY	NN	NY+NN
Total	YY+NY	YN+NN	YY+YN+NY+NN

5.1.3 Accumulating the forecast/observation pairs

In general, the forecast/observation pairs were accumulated over the 20-minute window that surrounds the forecast valid time. Thus, each 5-minute NCWD file that fell within the 20-minute window was used to create the statistics for that valid time. An example of the results computed for one issue/lead time for the CCFP is shown along the left-hand side of Figs. 1a and b. By accumulating the pairs within the 20-minute window for all forecast issue and lead times, the statistics can be computed for the entire evaluation period.

5.1.4 Computing coverage for CCFP

In addition to the *Yes/No* statistics, the actual (observed) coverage for each CCFP area was computed using the NCWD and displayed on the graphics as shown by the example in Fig. 1b. The procedure allows the forecasters to directly compare a computed coverage to the forecast coverage for each forecast issued. The actual coverage percentage for a given CCFP shape was computed by dividing the area covered by the NCWD by the total area of the particular CCFP shape. The area of the NCWD was computed by summing the area of all 40-km *Yes* NCWD boxes that fell within the forecast area. The total forecast area was computed by summing the area of all of the forecast 40-km boxes that fell within the forecast area. Each forecast area was handled independently, including those forecasts that fell within the domain of another forecast shape. Additional investigation and greater understanding of the coverage and probability categories is required in order to develop a verification methodology that would incorporate the forecast coverage criterion directly in the contingency table evaluations. Thorough testing of methodologies also would be required.

5.1.5 Verifying the CCFP attributes

Currently, the cloud tops, movement, and growth attributes from the CCFP are excluded from the verification scheme. Moreover, statistics are stratified by forecast coverage and then by probability, since the link between these two attributes is unclear and is difficult to capture in the verification scheme. Statistics are computed for each coverage and probability threshold (listed in Table 2) and then combined over all coverage thresholds and all probability thresholds. The overall statistics are summarized in the report. The reader is encouraged to access the Web site for additional statistics pertaining to a specific coverage or probability threshold.

5.2. Statistical measures

The verification methods used in this study are based on standard verification concepts (Brown et al. 1997; Murphy and Winkler 1987). The methods were developed by the Quality Assessment Group of the FAA Aviation Forecast and Quality Assessment PDT, and the Convective Weather PDT (Sankey 1997). To ensure that the study was complete and fair, statistics were generated using various verification techniques.

Table 4 lists the verification statistics used in this evaluation. The PODy and PODn are estimates of the proportion of *Yes* and *No* observations that were correctly forecasted, respectively (Brown et al. 1999; Brown et al. 1997). FAR represents the proportion of *Yes* forecasts that were incorrect. The Bias is the ratio of the number of *Yes* forecasts to the number of *Yes* observations and is a measure of over and underforecasting. The Critical Success Index (CSI), also known as the Threat Score, is the proportion of hits that were either forecast or observed. The True Skill Statistic (TSS) (Doswell et al. 1990) is a measure of the ability of the forecasts to discriminate between *Yes* and *No* observations, and is also known as Hanssen-Kuipers discrimination statistic (Wilks 1995). The Heidke Skill Score (HSS) is the percent correct, corrected for the number expected to be correct by chance. The Gilbert Skill Score (GSS; Schaefer 1990), also known as the Equitable Threat Score, is the CSI corrected for the number of hits expected by chance. The % Area is the percentage of the area of the forecast domain where convection is forecast to occur (Brown et al. 1997). Area Efficiency is the ratio of PODy to % Area. Most of the results presented here concern PODy, FAR, Bias, CSI, and % Area, but other statistics are included in the Web-based results.

6. Results

Overall results for the C-SIGMETs are summarized in Section 6.1. Section 6.2 summarizes the overall results for the CCFP.

6.1. Results for C-SIGMET

6.1.1 Overall comparisons

Overall results for the C-SIGMETs from 1 April-31 October 2001 are shown in Table 5. The numbers listed in Table 5 were generated for each forecast lead time by combining the pairs for all issue times over the 214-day period. During the period, forecasters issued 27,239 C-SIGMET polygons.

Table 4. Verification statistics used in this study

Statistic	*Definition*	*Description*
PODy	YY/(YY+NY)	Probability of Detection of "Yes" observations
PODn	NN/(YN+NN)	Probability of Detection of "No" observations
FAR	YN/(YY+YN)	False Alarm Ratio
CSI	YY/(YY+NY+YN)	Critical Success Index
Bias	(YY+YN)/(YY+NY)	Forecast Bias
TSS	PODy + PODn – 1	True Skill Statistic
Heidke	[(YY+NN)-C1]/(N-C1), where N=YY+NY+NY+NN C1=[(YY+YN)(YY+NY) + (NY+NN)(YN+NN)] / N	Heidke Skill Score
Gilbert	(YY-C2)/[(YY-C2)+YN+NY], where C2=(YY+YN)(YY+NY)/N	Gilbert Skill Score
% Area	(Forecast Area) / (Total Area) x 100	% of the area of the forecast domain where convection is expected to occur
Area efficiency	(PODy x 100) / % Area	PODy (x 100) per unit % Area

Comparisons between the statistics shown in Table 5 for the various forecast lead times indicate that the C-SIGMET polygon is best at capturing the convective activity at the 0-h lead, which should be the case since the criteria for generating the 0-h polygon is based the current reports of radar and lighting observations. Interestingly, the PODy drops by nearly half by the end of the 2-h lead, with the FAR increasing from a value of 0.69 to 0.81 during that same time. These results indicate that after 2 h, the convective activity captured by the C-SIGMET polygon is reduced, leaving areas inside the forecast area that may in fact be clear of convection. For future comparisons, the values presented in Table 5 can be used to baseline the skill for the C-SIGMETs. Comparisons of these values will aid decision-makers in evaluating changes or improvements in overall forecast skill of the C-SIGMET.

Table 5. Verification results for the C-SIGMET forecasts for 214 days from 1 April–31 October 2001 verified by the NCWD for each lead time where all issue times were combined. A baseline for forecast skill.

Forecast	Lead Time	PODy	PODn	FAR	CSI	TSS	HSS	GSS	Bias	%Area	Area Eff.
C-SIGMET	0	0.44	0.99	0.69	0.22	0.42	0.36	0.22	1.4	1.8	24
C-SIGMET	1	0.35	0.99	0.75	0.17	0.33	0.28	0.16	1.4	1.7	20
C-SIGMET	2	0.25	0.99	0.81	0.12	0.24	0.20	0.11	1.3	1.7	14

The character of the convection inside the forecast area impacts the minimum and maximum bounds that depict the statistical measures. Recall from Section 2, that the 1- and 2-h C-SIGMET forecasts are simply the 0-h C-SIGMET polygon moved to a new location. Moreover, the 0-h C-SIGMET polygon represents the forecaster's best ability to capture and group areas of active convective activity that are a threat to aviation. As a result, the character of the C-SIGMET polygon is such that the forecast area contains groupings of convective and nonconvective activity, as shown in Figs. 1a and b. This has an important impact on the range of the verification scores that can ultimately be computed. Perfect scores such as PODy = 1.0 and FAR = 0.0 are unattainable.

Therefore, we propose that the scores for the 0-h C-SIGMET be used as an upper bound for judging the quality of other convective forecasts, such as the CCFP. Therefore, as shown in Table 5, the results generated for CCFP will be compared in Section 6.2 to these upper bounds of the forecast skill: PODy = 0.44, FAR = 0.69, HSS = 0.36, GSS = 0.22, Bias = 1.4, and %Area = 1.8.

These "upper bounds" can also be used to compare to the 1- and 2-h C-SIGMET results. In particular, use of these upper bounds indicates that

- The 1- and 2-h PODy values in Table 5 are 80 and 57% of the upper bound for PODy;
- The FAR values for the 1- and 2-h C-SIGMET results are 81% and 61% as good as the FAR values for the 0-h C-SIGMETs;
- The CSI values are 77% and 55% of the upper bound for CSI;
- The HSS values are 78% and 56% as big as the maximum; and
- The GSS values are 73% and 50% of the maximum values.

6.1.2 Improvements in the C-SIGMET

When trying to show improvements in forecast skill, the year-to-year variability in the weather poses a complicating factor. In addition, improvements in skill can only be realized when the values of at least two statistics improve over time. Therefore, with these rules in mind, the skill of the C-SIGMET was evaluated by comparing a subset of results covering the four-month period extending from 1 April-31 July of 2001 and 2002. The results are presented in Figs. 2 a-d and Table 6.

Combining all forecast/observation pairs for all issue times per month for the C-SIGMET 0-h lead created the statistics shown in Figs. 2 a-d. The forecasts were compared over the 4-month period using verification methods that remained consistent. As shown in Figs. 2 a-d, substantial improvements in the quality of the C-SIGMET forecasts in 2002 are evident. For instance, the PODy increased nearly 0.1 each month during 2002, the values of FAR during 2002 were generally less than or equal to the 2001 values, and the CSI value was larger each month during 2002 than those computed in 2001. During May and July 2002, the Bias remained lower than the value computed in 2001.

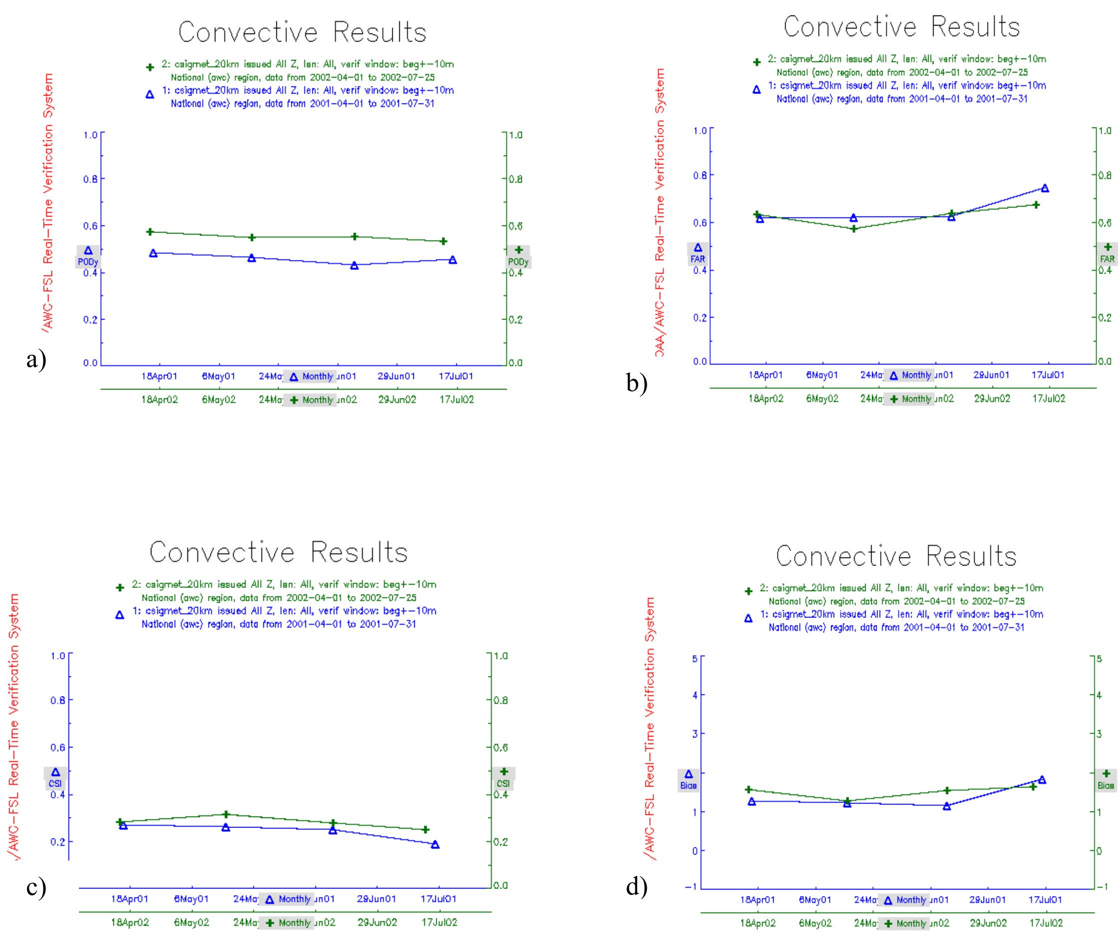


Fig. 2. Time series plots of monthly PODy (a), FAR (b), CSI (c), and Bias (d) between 1 April–31 July 2001 (triangle) and 2002 ('+'). Values created by combing pairs for all issues times per month for the 0-h lead.

The overall statistics for the spring/early summer seasons of 2001 and 2002 for the C-SIGMETs are shown in Table 6. For all forecast leads, the statistics for PODy, PODn, FAR, and CSI are best for the C-SIGMET forecasts issued in 2002. The largest increases occurred at the 0-h lead with the PODy value increasing from 0.46 to 0.55 *and* the FAR value decreasing from 0.68 to 0.64, although improvement in forecast skill was noted for all forecast lead times. These improvements in forecast skill are quite possibly due to the standardization of the procedures and approaches for issuing a C-SIGMET that were implemented at AWC at the end of the 2001 season.

Table 6. Verification results for the C-SIGMET forecasts for the spring and early summer season from 1 April–31 July 2001 and 2002 verified by the NCWD for all forecast issue and lead times combined.

Forecast	Lead Time	PODy	PODn	FAR	CSI	Bias	%Area
2001							
C-SIGMET	0	0.46	0.99	0.68	0.23	1.4	2
C-SIGMET	1	0.36	0.98	0.75	0.18	1.4	2
C-SIGMET	2	0.26	0.98	0.81	0.12	1.4	2
2002							
C-SIGMET	0	0.55	0.99	0.64	0.28	1.5	2
C-SIGMET	1	0.43	0.99	0.72	0.20	1.5	2
C-SIGMET	2	0.31	0.98	0.80	0.14	1.5	2

6.2. Results for CCFP

6.2.1 Overall Comparisons

Combining all forecast/observation pairs per week for all issue and lead times from 1 April – 31 October 2001 produced the statistics shown in Fig. 3. The largest variability in the weekly statistical results occurred during April, May, and October when the convective activity is at a minimum. During the summer months when convection is at its peak, little or no variability in the scores (i.e., PODy, FAR, and CSI values) is apparent. Generally, PODy values remain around 0.3, FAR values at 0.8, and CSI values at or below 0.2

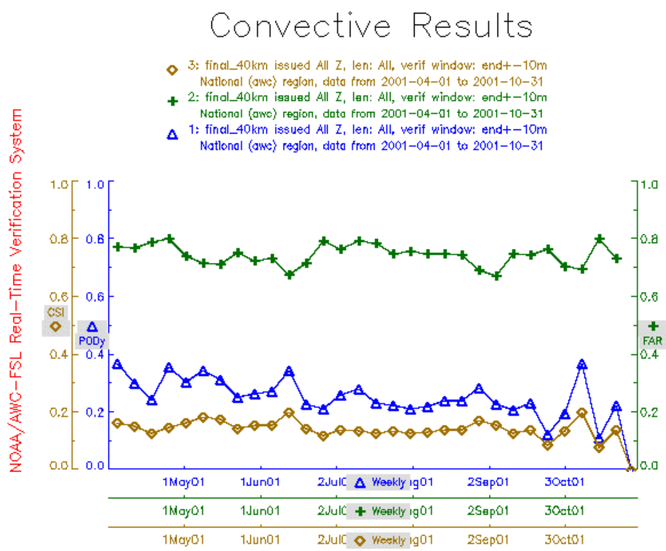


Fig. 3. Time series plots of weekly PODy (triangle), FAR (+), and CSI (diamond) for all CCFP issue and lead times combined from 1 April–31 October 2001.

Overall statistics are presented in Table 7 and were computed for each forecast lead time by combining the forecast-observation pairs for all issue times over the 214-day period. During the period forecasters issued a total of 8,139 CCFP polygons during 2001. Comparisons between the different forecast lead times indicate that the CCFP polygon is best at capturing the convective activity at the 2-h lead. Interestingly by the 6-h lead, the PODy drops from 0.30 to 0.21, the FAR increases from a value of 0.69 to 0.79, and the Bias remains consistent at 1.0, indicating no over or underforecasting throughout the 6-h forecast period.

The overall values of the skill scores for 2000 and 2001 are listed at the bottom of the table. These scores represent the overall statistics when all lead and forecast times are combined. Interestingly, as measured by the PODy, CSI, and HSS, the skill of the CCFP was slightly better in 2000 than in 2001. However, the percentage of forecast area generally covered by a CCFP forecast in 2001 improved by decreasing from 3.2% to 2.6%. The FAR also decreased slightly in 2001. These results suggest that although the forecast area didn't capture quite as much of the convective activity in 2001 as it did in 2000, the false alarms were reduced and more area (not covered by a forecast) was made available for moving air traffic throughout the National Air Space (NAS) system.

Table 7. Verification results for the CCFP for 214 days from 1 April–31 October 2001 verified by the NCWD, for all height, probability, and coverage categories combined and for each lead-time where all issued times were combined. The 2000 CCFP results were recomputed excluding the 0700 and 1100 UTC issue times and the 0300 UTC, 6-h lead so that 2000 and 2001 results could be directly compared.

Forecast	Lead Time	Number of Forecasts Issued (Areas and Lines)	PODy	PODn	FAR	CSI	TSS	HSS	GSS	Bias	%Area
CCFP	2	3203	0.30	0.98	0.69	0.18	0.28	0.28	0.16	0.97	2.6
CCFP	4	2864	0.25	0.98	0.76	0.14	0.23	0.22	0.13	1.00	2.6
CCFP	6	2415	0.21	0.98	0.79	0.12	0.19	0.19	0.10	0.98	2.7
Total CCFP2001	All	8139	0.25	0.98	0.74	0.14	0.23	0.23	0.13	0.99	2.6
Total CCFP2000	All		0.28	0.97	0.75	0.15	0.25	0.24	0.13	1.1	3.2

6.2.2 Meaning Behind the CCFP Statistics

To give meaning to the scores listed in Table 7, a comparison between the CCFP and the 0-h C-SIGMETs is presented.

Recall from Section 2, the CCFP polygon is generated through a collaborative process where the resultant forecast area defines the convective activity that is a threat to aviation over a 2-, 4-, and 6-h period. Much like the C-SIGMET, the character of the CCFP polygon is such that the forecast area contains groupings of convective cells and groupings of nonconvective activity as was shown in Fig. 1b. Unlike the C-SIGMET, the character of the convection encompassed by the CCFP area is defined by the "forecast coverage" attribute. Nevertheless, the character of the convection within a forecast area impacts the range of verification scores that are computed, and such ideal scores as PODy = 1.0 and FAR = 0.0 are unattainable. Moreover, if the scores presented in Table 7 were compared to these ideal scores, the forecast skill would be deemed as "bad." However, when the scores listed in Table 7 are compared to the scores generated for the 0-h C-SIGMETs, the skill of the CCFP becomes more acceptable to the users.

Fig. 4 shows a comparison between the CCFP results and those generated for the 0-h C-SIGMETs, which (as described earlier) will be used as an upper bound to the statistical results. These results are presented so that the reader begins to understand the skill of the CCFP as it relates to other similar forecasts of convection. The statistics for the C-SIGMET 0-h lead are used for comparison since that forecast closely represents the current convective activity and is a forecast that is similar to the CCFP. The results for PODy for the 2-h CCFP is 68% of the PODy computed for the C-SIGMETs. The FAR value for the CCFP at a 2-h lead is equivalent to the FAR computed for the C-SIGMET at a 0-h lead. At longer forecast leads (i.e., 4- and 6-h) for the CCFP, the FAR values are

77% and 68% as good as that computed for the 0-h C-SIGMETs. The CSI for the 2-h CCFP is 81% of the CSI computed for the C-SIGMETs. Therefore, the skill of the CCFP at a 2-h lead is comparable to the 0-h C-SIGMETs. However, the primary difference between the two forecasts is the percentage of the total forecast domain covered by the forecasts. For instance, the percentage of forecast area covered by the C-SIGMET is, on average, only one-third the size of the CCFP. These larger areas are partly due to the greater uncertainty associated with longer-lead CCFP forecasts, but also may be associated with the intended use of the CCFP (i.e., traffic flow management) in contrast to the intended use of the C-SIGMETs (i.e., warning pilots of dangerous areas). That is, the forecasters may be creating larger forecast areas simply in anticipation of the decisions that will be made by users of the forecasts. The statistics presented here provide further support for use of a 40-km scale for verification of the CCFP forecasts.

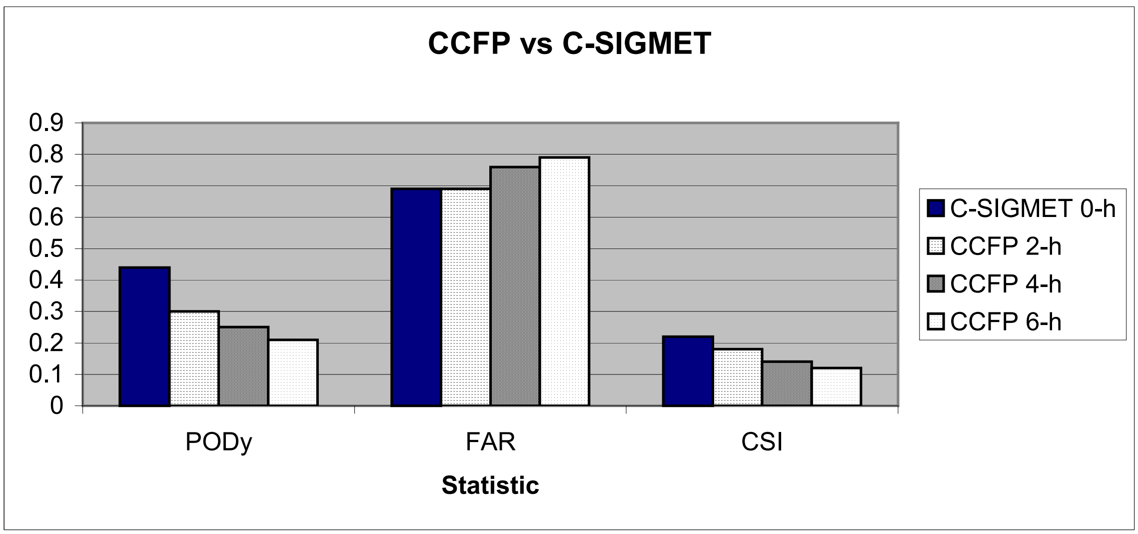


Fig. 4. Histogram of PODy, FAR, and CSI values for the 0-h C-SIGMET (solid), the 2-h CCFP (dashed), the 4-h CCFP (hatched), and the 6-h CCFP (dotted). All forecast issue and lead times for 1 April–31 October 2001 were combined to create these statistics**.**

6.2.3 Area coverage and probability

The frequency of use of the various coverage and probability categories are shown in Table 8. This table indicates that the Low coverage category was used almost all the time. The Medium coverage category was used only 8% of the time, and the High coverage category was used less than 1% of the time. This result might be interpreted as suggesting that the Low coverage category is used too frequently; however, the information in Table 8 is inadequate to determine if that is the case (see results later in this section). However, the fact that 92% of the CCFPs forecasted Low coverage (25-49%) is consistent with the overall FAR value of 0.75 (Table 7). Thus, the CCFP FAR values are not at all unreasonable.

In contrast to the coverage categories, the Low and Medium Probability categories were used with approximately equal frequency. The High Probability category was used quite infrequently (only 3% of the time).

Although the coverage and probability categories were not directly considered in the binary verification analyses, a supplemental analysis was done to determine the accuracy of the forecasted coverage categories. In this analysis, the actual coverage values for each CCFP shape are compared to (a) the forecast coverage category and (b) the forecast probability category.

Table 8. Actual frequency of use of the various coverage and probability categories during the 2001 evaluation period (1 April–31 October 2001).

Coverage	*Probability*			Total
	LOW	**MEDIUM**	**HIGH**	
LOW	51%	40%	1%	92%
MEDIUM	1%	6%	1%	8%
HIGH	--	--	--	<1%
TOTAL	51%	46%	3%	(n=8,433)

Figure 5 shows distributions of actual coverage for each forecast coverage category. The distributions of actual coverage, represented by the box-and-whisker plots are a convenient way to compare two or more distributions. The "box" part of the plot is enclosed by the 75th and 25th percentiles of the distribution, with the line inside representing the median (i.e., 50th percentile) value; thus, the box represents the middle 50% of the observed distribution of values. The "notches" on the boxes represent 95% confidence interval bounds for the median. The ends of the "whiskers" extending above and below the box represent the 95th and 5th percentiles of the distributions, respectively (i.e., 5% of the observations are larger than the end of the upper whisker, and 5% are smaller than the end of the lower whisker). Points above and below the whiskers are the extreme observations in the upper and lower 5% of the observations.

The plots in Fig. 5 indicate that the areal coverage forecasts were quite reliable, overall. In fact, the box parts of the figure indicate that the median coverage values were 22%, 42%, and 62%, for the Low, Medium, and High coverage forecasts, respectively. Moreover, the medians for the three groups are statistically significantly different from each other, as indicated by the positions of the notches. This figure indicates that Low coverage forecasts are generally associated with low coverage events, Medium coverage forecasts are generally associated with medium coverage events, and High coverage forecasts are generally associated with high coverage events. Moreover, it is clear that most of the CCFP shapes are associated with relatively low actual coverage: The median actual coverage for all forecasts is 25%; the 75th percentile is 40%; and the 90th percentile is 55%. The results for individual lead times (not shown) are consistent with the overall results shown in Fig. 5.

The distributions of actual coverage as a function of the forecast probability category are shown in Fig. 6. This figure indicates that the relationship between actual coverage and probability is not nearly as strong as the relationship to predicted coverage. The median values of actual coverage increase from 19% for Low probability forecasts to 28% for Medium probability forecasts, to 42% for High probability forecasts.

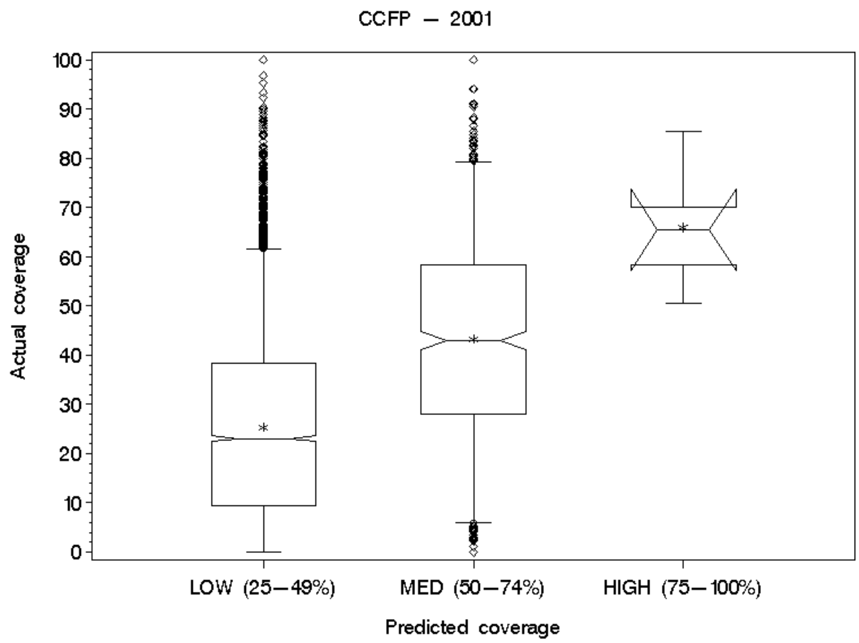


Fig. 5. Box plots showing distributions of actual coverage as a function of predicted coverage, for CCFP forecasts issued during the 2001 evaluation (1April–31 October 2001).

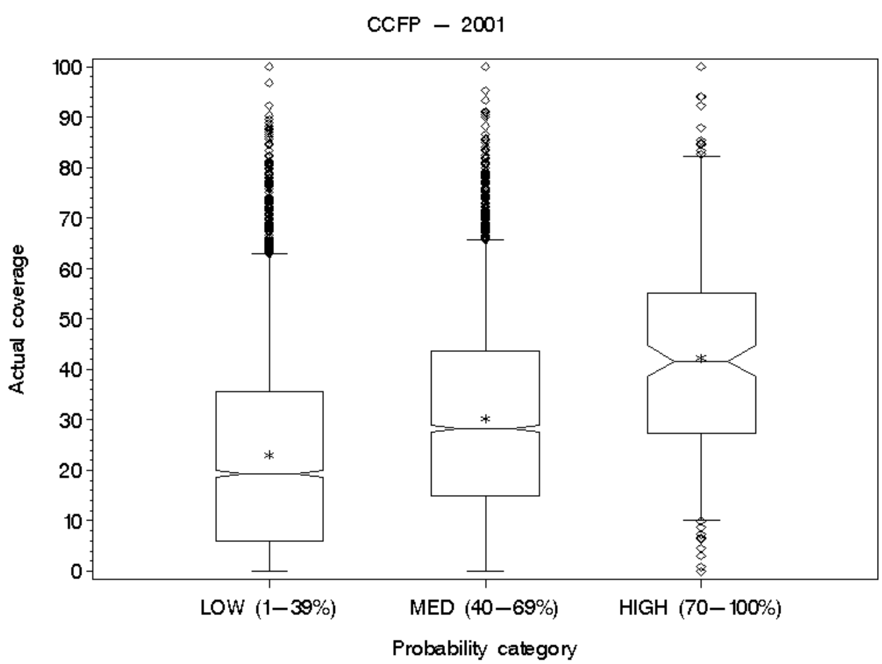


Fig. 6. Actual coverage as a function of forecast probability.

6.2.4 Understanding the differences between the 2-h CCFP and 2-h C-SIGMET

The CCFP and the C-SIGMET forecasts are both issued by NWS forecasters for areas of convective activity. Both include a 2-h forecast and both are human derived, although there are slight differences in the manner and rules for which they are derived. The CCFP is designed to provide guidance for traffic flow management and the C-SIGMET is designed to warn general aviation pilots of dangerous areas that are associated with convective activity. With these two differing goals, the aviation community often misuses these forecasts. Therefore, in this section, the two forecasts are compared so that differences and similarities between the two forecasts are revealed. However, since the CCFP and the C-SIGMETs are at different scales, further interpretation of the forecast quality is not advised.

Three sets of displays showing the C-SIGMETs and CCFP forecasts overlaid with the NCWD are presented in Figs. 7-9. The displays are presented to show the character of the two forecasts at the 2-h lead. The reader must keep in mind that the C-SIGMET is issued at time=0 and is allowed to move at a predefined speed for 2 hours. The CCFP, on the other hand, is developed through a collaborative process and is issued with the location of where convection will occur in two hours.

The weather depicted in Figs. 7 and 8 show well-organized convection generally grouped nicely in small-localized areas. The weather shown in Fig. 9 is wide spread and is unorganized and numerous. By investigating these three cases, several differences/similarities between the C-SIGMET and CCFP become apparent. First, when convection is nicely organized and grouped, the CCFP and the C-SIGMET seem to be similar in size, and location as shown in Figs. 7 and 8. However, when convection is wide spread and numerous, the size of the CCFP gets larger than the C-SIGMET by trying to encompass more of the convective activity in larger forecast areas rather than breaking down the forecast areas into smaller more numerous areas. In addition to these observations, the CCFP includes attributes for coverage and probability that are not a part of the C-SIGMET forecast.

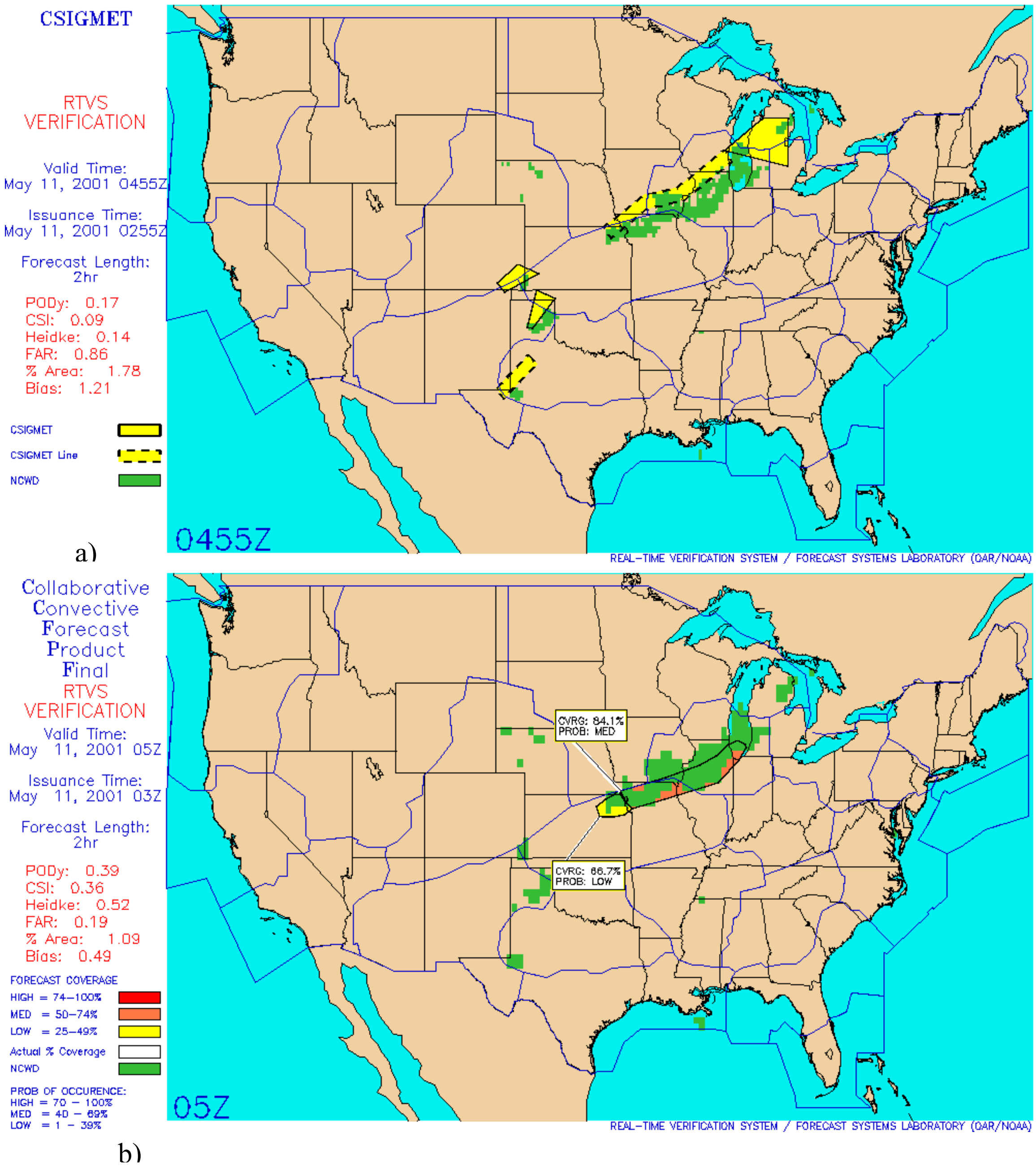
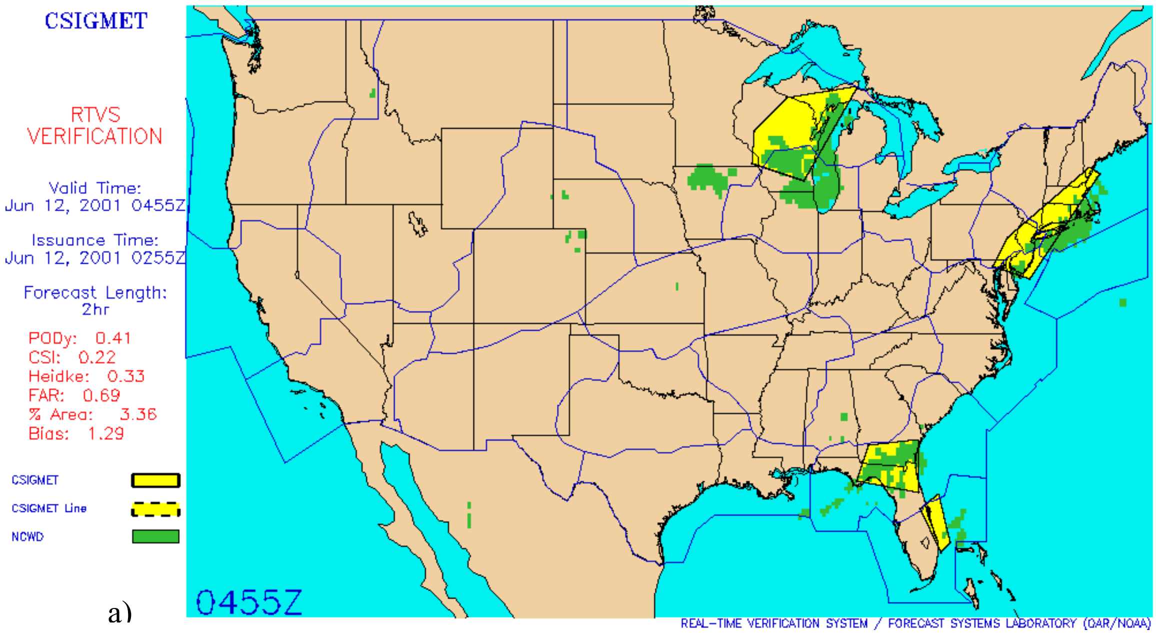
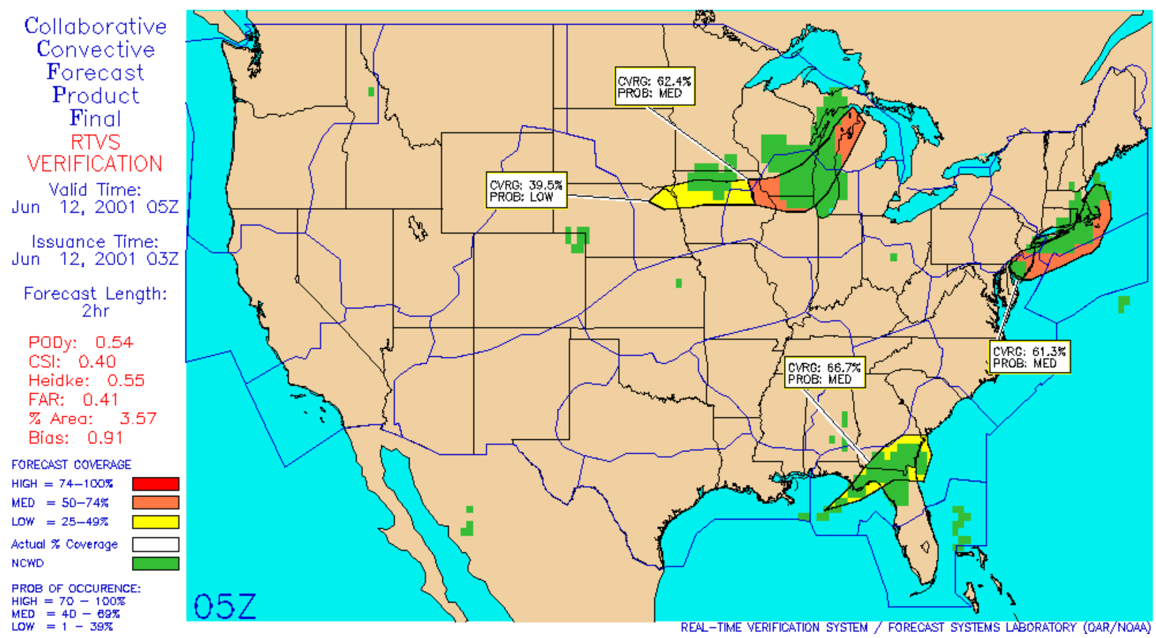


Fig. 7. Forecast maps depicting the C-SIGMET (a) and CCFP (b) forecasts (circles) and the NCWD (gray squares) 2-h lead issued on 11 May 2001, valid at 0500 UTC. White boxes on b) indicated computed coverage. Statistics are listed on the left side of display.

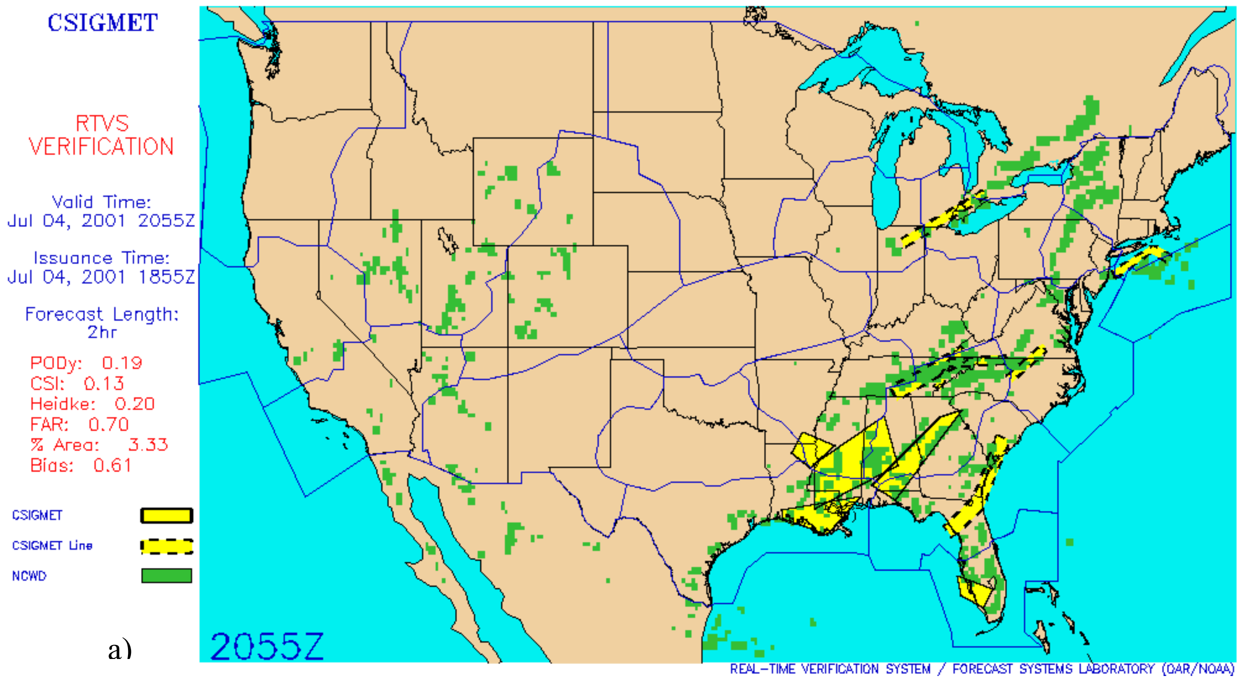


a)

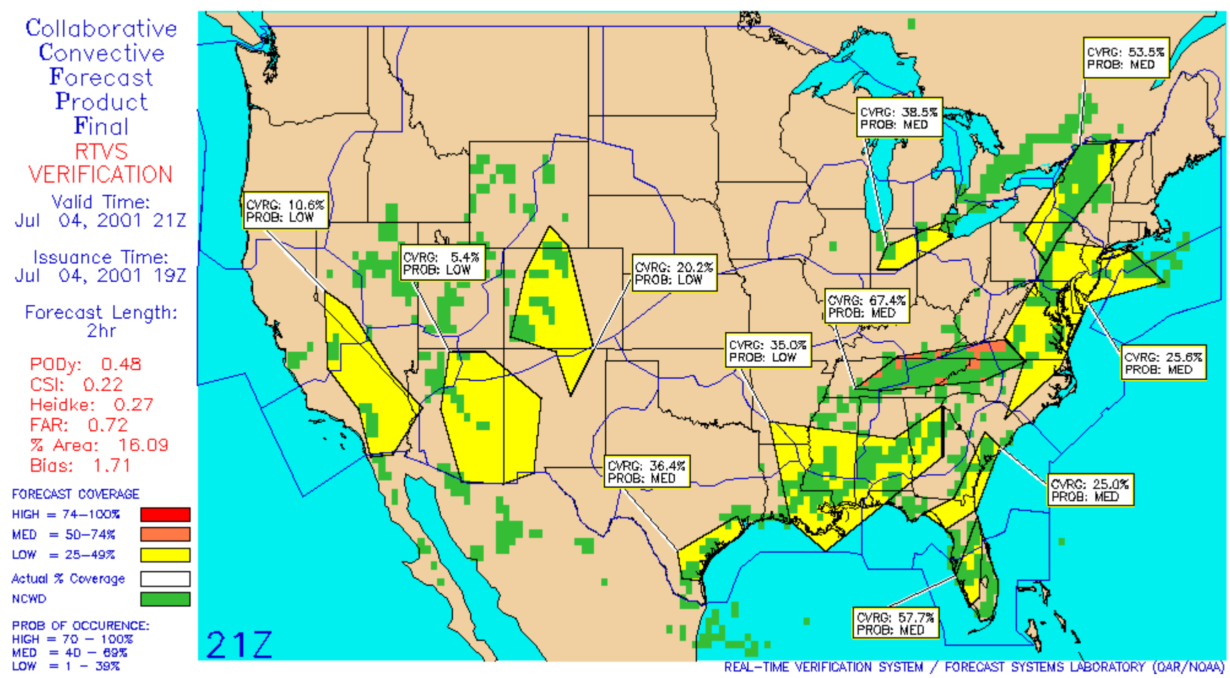


b)

Fig. 8. Same as in Fig. 7, except for 2-h lead issued on 12 June 2001, valid at 0500 UTC.



a)



b)

Fig. 9. Same as in Fig.7, except for 2-h lead issued on 4 July 2001, valid at 2100 UTC.

Overall statistics by forecast lead for the C-SIGMET and CCFP are presented in Fig. 10 a-d. As noted earlier, the best forecast skill occurs at shorter lead times for both the CCFP and the C-SIGMET. When comparing the skill of the 2-h lead, the skill of the CCFP is somewhat better than that of the C-SIGMET as indicated by a larger PODy and CSI, and smaller FAR and Bias, as shown in Fig 10 a-d. However, one must keep in mind that the forecasts are mapped to a different sized grid and the C-SIGMET area is geared toward providing guidance at the 0- and 1-h leads.

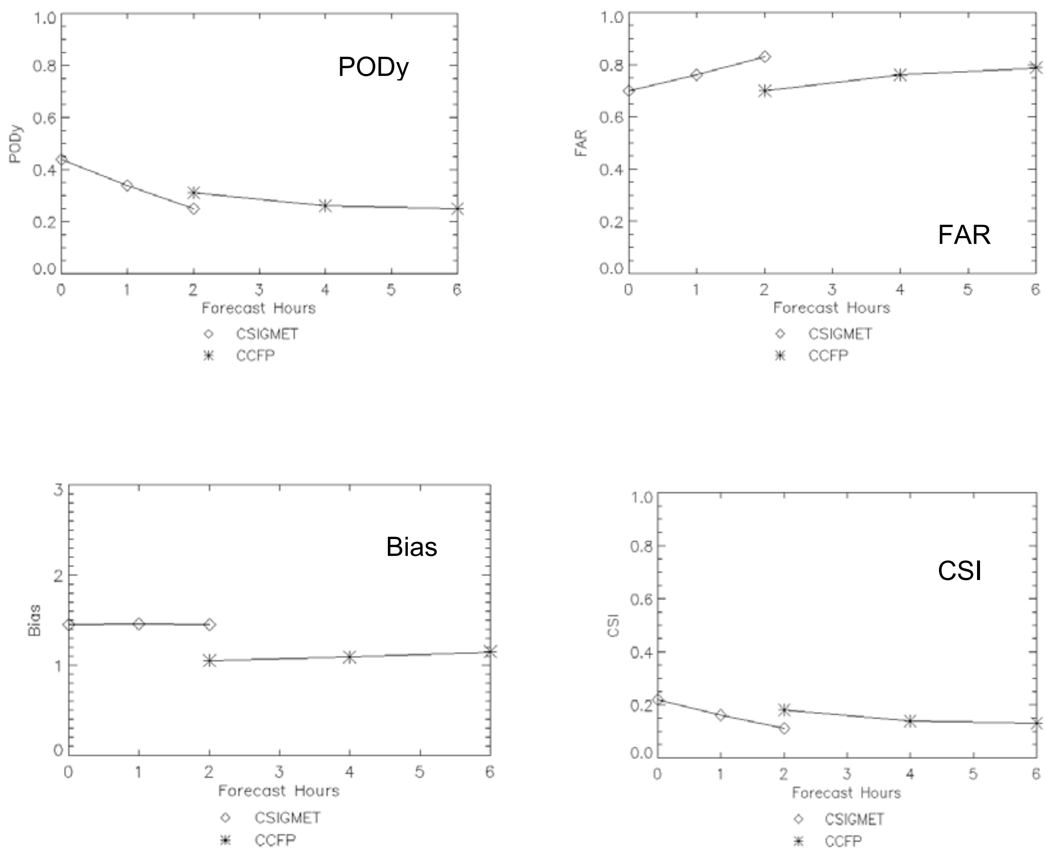


Figure 10. Overall PODy (a), FAR (b), Bias (c), CSI (d), for the 2001 CCFP and C-SIGMETs as a function of lead time.

7. Summary and conclusions

An evaluation of the C-SIGMET and CCFP forecasts was presented in this report. Based on current verification approaches, a statistical baseline was established for the C-SIGMETs (Table 6). The baseline was used to show that the forecast skill of C-SIGMET forecasts improved between 2001 and 2002. This improvement possibly resulted from the standardization of procedures and approaches for issuing C-SIGMETs that was implemented at AWC at the end of the 2001 season. In addition, the percentage of area within the NAS generally covered by the CCFP forecast was reduced in 2001 when compare to 2000. However, although the CCFP warned smaller areas in 2001, it was also less efficient at capturing the convection than in 2000. When convective activity is wide spread and numerous, the size of the CCFP often gets much larger than the C-SIGMETs by trying to encompass more of the convection in a smaller number of forecast areas. The two forecasts seem quite similar in character when the convection is well organized and located in a limited area. However, these larger areas of the CCFP are partly due to the greater uncertainty associated with the longer-lead forecasts, but also may be associated with the intended use of the CCFP (i.e., traffic flow management) in contrast to the intended use of the C-SIGMETs (i.e., warning pilots of dangerous areas). That is, the forecasters may be creating larger forecast areas simply in anticipation of the decisions that will be made by users of the forecasts. The statistics presented here provide further support for use of a 40-km scale for verification of the CCFP forecasts.

When 0-h C-SIGMET statistics were used as the upper bounds for the CCFP, skill of the CCFP appeared to dramatically improve (with relative PODy skill values of 68%, relative FAR skill values ranging from 67-100%, and relative CSI skill values reaching 81%).

An analysis of the actual convective coverage associated with each CCFP shape indicates that the forecast coverage categories are actually quite reliable: Low coverage forecasts are generally associated with low actual coverage; Medium coverage forecasts are generally associated with medium actual coverage; and High coverage forecasts are generally associated with high actual coverage. These results suggest that the coverage forecasts may provide a good representation of the probability of convection in an area, or at any point in a given area.

Future work includes continuing the evaluation of the CCFP through the 2002 convective season, investigating enhancements to the verification methods that allow for detailed analysis of the forecast location and orientation errors, and adding the ability to verify improvements that will be associated with the 2002 CCFP forecasts.

Acknowledgments

This research is in response to requirements and funding by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government. We would like to thank Nita Fullerton for her review of this work and to the AWC staff for contributing to this evaluation.

References

Brown, B.G., J.L. Mahoney, R. Bullock, J. Henderson, and T.L. Kane, 1999: Turbulence Algorithm Intercomparison: 1998-1999 Initial Results. FAA Turbulence Product Development Team Report to FAA Aviation Weather Research Program (available from the author at Research Applications Program, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000).

Brown, B.G., G. Thompson, R.T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of in-flight icing algorithms. Part II: Statistical verification results. *Wea. Forec*asting, **12,** 890-914.

Doswell, C.A., R.Davies Jones, and David L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576-585.

Hudson, H. R. and F. P. Foss, 2002: The Collaborative Convective Forecast Product from the Aviation Weather Center's Perspective. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Oregon, WA, Amer. Meteor. Soc., 73-76.

Kulesa, G.J., P.J. Kirchoffer, D.J. Pace, W.L. Fellner, J.E. Sheets, and V.S. Travers, 2002: New weather products developed by the Federal Aviation Administration's Aviation Weather Research Program. *Preprints, 10th Conference on Aviation, Range, and Aerospace Meteorology*, Portland, OR, Amer. Meteor. Soc. (Boston), 18-19.

Mahoney, J.L., B.G. Brown, J.E. Hart, C. Fischer, 2002a: Using verification techniques to evaluate differences among convective forecasts. *Preprints, 16th Conference on Probability and Statistics in the Atmospheric Sciences*, Orlando, FL, Amer. Meteor. Soc. (Boston), 12-19.

Mahoney, J.L, J. K. Henderson, B.G. Brown, J.E. Hart, A. Loughe, C. Fischer, and B. Sigren, 2002b: The real-time verification system (RTVS) and its application to aviation weather forecasts. *Preprints,* 10th *Conference on Aviation, Range, and Aerospace Meteorology*, OR, Amer. Meteor. Soc. (Boston), 323-326.

Mahoney, J.L., B.G. Brown, R. Bullock, J.E. Hart, C. Wallace: 2001: The second evaluation of the Collaborative Convective Forecast Product (CCFP): Spring and Summer 2000. Report submitted to AUA-100 (available from the author at mahoney@fsl.noaa.gov).

Mahoney, J.L., B.G. Brown, C. Mueller, and J.E. Hart, 2000a: Statistical Verification Results for the Collaborative Convective Forecast Product. NOAA Technical Report OAR 457-FSL 6 (available from the author, 325 Broadway, Boulder, CO 80305-3328).

Mahoney, J.L., B.G. Brown, C. Mueller, and J.E. Hart, 2000b: Convective intercomparison exercise: Baseline statistical results. *Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology*, Orlando, Fl., Amer. Meteor. Soc., 403-408.

Mahoney, J.L., J.K. Henderson, and P.A. Miller, 1997: A Description of the Forecast Systems Laboratory's Real-Time Verification System (RTVS). *Preprints, 7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, Amer. Meteor. Soc., J26-J31.

Mueller, C.K., C.B. Fidalego, D.W. McCann, D. Meganhart, N. Rehak, and T. Carty, 1999: National Convective Weather Forecast Product. *Preprints, 8th Conference on Aviation Range, and Aerospace Meteorology*, Amer. Meteor. Soc. (Boston), 230-234.

Murphy, A.H. and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

NWS, 1991: National Weather Service Operations Manual, D-22. National Weather Service (available at Website http://www.nws.noaa.gov).

Orville, R.E., 1991: Lightning ground flash density in the contiguous United States-1989. *Mon. Wea. Rev.*, **119**, 573-577.

Phaneuf, M. W. and D. Nestoros, 1999: Collaborative convective forecast product: Evaluation for 1999 (available from the author at AvMet Applications, Inc.).

Sankey, D., K.M. Leonard, W. Fellner, D.J., Pace, K.L. Van Sickle, 1997: Strategy and Direction of the Federal Aviation Administration's Aviation Weather Research Program. *Preprints, 7th Conference on Aviation, Range, and Aerospace Meteorology*, Long Beach, Amer. Meteor. Soc., 7-10.

Schaefer, J.T., 1990: The Critical Success Index as an indicator of warning skill. *Wea. Forec*asting, **5**, 570-575.

Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.